

自然な音声対話における間投詞・ポーズ・ 言い直しの出現パターンと音響的性質*

中川 聖一^{*1} 小林 聡^{*1,†}

(1994年5月12日受付)

【要旨】 人間の自然な発話には間投詞や言い直しが含まれる。特に、間投詞に関しては、1文に1個以上の割合で出現し、また言い直しに関しても10文に1~2回程度の頻度で出現している。更に言い直しは、1文が長くなるほど現れ易い傾向がある。そのため、自然な発話を扱おうとすると、間投詞や言い直しに対処する必要がある。本論文では、まず間投詞語彙の出現位置と出現頻度及び音響的性質を調べ、その統計情報に基づく音声認識法を述べる。次に、ポーズと間投詞及び言い直しの出現の関連を述べる。また、言い直しの形態と、音響的性質について述べる。

キーワード 自然な発話, 間投詞, 言い直し, ポーズ, 音声認識

Spontaneous speech, Interjection, Repair, Pause, Speech recognition

1. はじめに

実用的な音声対話システムの実現にあたっては、より自然な発話を受理できなければならない。つまり、人間同士が自然言語で対話を行うように、自然な対話が可能になるのが望ましい。しかし、音声による自然な対話を目指した場合、キーボードからの自然言語による対話や、整った話し言葉による対話などの場合では問題とならなかった事柄が新たな問題となる。通常、人間同士での音声による対話を考えてみると、そこには様々な ill-formedness (非文) が現れる。このような問題を解決しなければ、音声による自然な対話の実現は難しいと思われる。

自然な対話に含まれる ill-formedness としては、助詞落ち、倒置、間投詞、言い直しなどが考えられる。助詞落ち、倒置に関しては、文献2)においてその傾向や対処方法としてのヒューリスティックなルールが与えられている。また、間投詞に関しては、文献1)などにより、間投詞語彙の出現頻度が述べられ、文献14)においては不要語を一つの HMM にあてる方法での間投詞を含んだ文の認識実験が行われている。

言い直しは、文献15)によっては書き起こしテキストを用いてその形態の分類及び検出・訂正方法が提案されている。また、文献7)においては、形態の分類及び、発話内容のテキストを音声データと共に用いる方法で検出・訂正の実験が行われている。特に文献7)においては、テキストを用いての、単語の統語的類似性や、似たような単語の並び、更にはキーワードなどから言い直し部分の候補を得て、後に音響的な特徴を用いて、先に得られた候補をふるいにかけてという方法を用いて、同じ単語を繰り返す(繰り返しの間に挿入が行われる場合もある)形態の言い直しの90%を検出できている。また、C. Nakataniらは、ポーズの継続時間や、基本周波数、単語断片の存在などの、音響的特徴などを用い分類と回帰解析木(Classification and Regression Tree)を得る方法でやはり90%程度の検出を可能とする⁹⁾。言い直しの韻律的特徴に関しては、D. O'Shaughnessyなどによる文献8)において、言い直しはその対象となる部分よりも短く発話される傾向があり、言い直しの前にポーズが現れ易いなどの結果を述べている。

対話に現れるポーズに関しては、文献4), 5)などで、対話中の発声速度の変化や話題の変化による発話様態の変化などが調査されている。その結果、対話音声でのポーズ長は、朗読音声に比べて長くなり、その分布も複雑になるなどの特徴が得られている。また、保坂らは、自然な対話では意味理解を行い易いように、意味を表すのに適したところにポーズを置くと報告し、

* Phenomena and acoustic variation on interjections, pauses and repairs in spontaneous speech, by Seiichi Nakagawa and Satoshi Kobayashi.

*1 豊橋技術科学大学情報工学系
(問合先: 中川聖一 〒441 愛知県豊橋市天伯町字雲雀ヶ丘 1-1 豊橋技術科学大学情報工学系)

† 現在 静岡大学電子科学研究科

その特徴を利用した連続音声認識方法が提案されている¹²⁾。

しかし、間投詞の出現位置と出現パターン、音響的性質、ポーズとの関連などの分析はされておらず、また、言い直しに関しての日本語音声に対する音響的な特徴分析はなされていない。

本論文では、間投詞語彙の出現位置と出現頻度及び音響的変動を調べ、それに基づき、間投詞語彙を辞書登録する方法を用いた文の認識実験に関して述べる。また、ポーズと言い直し、間投詞との関係を述べる。言い直しに関しては、その形態と音響的変動を調べると共に、音声データのみによる検出実験について述べる。

2. 模擬対話データベース

今回の分析には、日本音響学会連続音声データベースの書き起こしテキストの一部を使用した。本研究では、間投詞と言い直しの分析に関しては、TUT 0001, TUT 0002, TUT 0003, TUT 0004, TUT 0005, TUT 0006, TUT 0008, TUT 0009, TUT 0010, TUT 0011 の計 10 対話を対象として行った。話者数などのデータを表-1 に示す。

なお、これらの音声データに関しては DAT に収録された、模擬対話の音声データを使用した。

これらのデータ中の、助詞落ち、倒置、間投詞、言い直しの出現頻度を表-2 に示す。

表-2 で分かるように、間投詞はその出現数が他のものに比べて非常に多く、1 文に 1 回以上現れていることになる。また、言い直しも、10 文に 1 回以上は現れ

ていることになり、自然な対話を処理する場合には、対処する必要がある。

3. 間投詞の特徴分析

3.1 間投詞語彙の出現頻度

間投詞は、単独の語彙が発声される場合と、複数の語彙が連続して発声される場合がある。複数の間投詞が連続して発声される割合を表-3 に示す。表-3 から分かるように、間投詞の 90% 近くは単独で発声されている。また、単独及び二つの間投詞が連続して発声される場合で、98% と間投詞出現のほぼすべてを占めている。

次に、間投詞が単独で発声されている場合、そこで用いられる語彙を表-4 に示す。なお、間投詞に含まれる長母音や、促音による無音区間の長さはまちまちであるため、それらの長母音や促音はここでは便宜上無視した上での集計である。この順位は、文献 3) に示されている長母音や促音を考えに入れた場合、つまり「えと」と「えーと」を別の語彙として考えた場合とほぼ同様の順位になっている。また、二つの間投詞が連続して発声される場合についての語彙を表-5 に示す。

表-3 間投詞の連続発声数

連続数	出現回数	[%]
1	1,044	88.1
2	117	9.9
3	22	1.9
4	2	0.2
間投詞総数		1,185 個

表-1 分析データベース

ドメイン	観光案内, 相談, 他の案内
対話数	10 対話
話者数	11 人
総文数	1,052 文
文節数	4,597 文節

表-2 Ill-formedness の出現数

Ill-formedness	出現回数	出現割合
助詞落ち	99 個	(0.094 回/1 文)
倒置	18 個	(0.018 回/1 文)
間投詞	1,185 個	(1.126 回/1 文)
言い直し	153 個	(0.145 回/1 文)

表-4 単独間投詞の内訳

間投詞	出現数	[%]	累積[%]
え	238	22.8	22.8
えと	230	22.0	44.8
あの	207	19.8	64.7
あ	162	15.5	80.2
ま	130	12.4	92.6
えとです	33	3.2	95.8
ん	9	0.9	96.6
そです	9	0.9	97.5
その	7	0.7	98.2
こ	5	0.5	98.7
あと	3	0.3	98.9
んと	2	0.2	99.1
は	2	0.2	99.3
その他	7	0.7	100.0

表-5 2連続間投詞の内訳

一つめ	個数	二つめ	個数
え	23	ま	42
あの	23	え	26
えと	19	あの	16
あ	18	えと	12
えとですね	14	あ	8
ま	9	そうですね	4
ん	3	その	2
は	2	ん	2
そうですね	2	その他	5
その他	4		

表-6 間投詞の語彙使用の個人差
(単独発声の場合)

話者 YM		話者 SK	
え	20	あの	28
ま	16	えと	25
あの	7	あ	18
えと	6	え	5
あ	4	ま	3
えとですね	2	ん	1
こ	1	あと	1
その	1	えとですね	1

表-4 の上位五つで、単独で発声される場合の 93%、また間投詞の出現総数の 82% を占めている。このことから、間投詞としては、少数の語彙が非常に多用されていることが分かる。

表-5 より、二つの間投詞が連続して発声される場合でも、用いられる語彙の上位に関しては、単独で発声される間投詞の語彙の上位とほぼ同じであることが分かる。

話者によって好んで用いる間投詞が異なることが考えられるが¹⁴⁾、表-6 を一例として考えると、語彙の順位は異なっているものの、やはり多用される語彙は共通となっている。

間投詞は、文の中での何等かの意味的な区切りとなるようなところに現れ易いのではないかと考え、次に示す四つ(前後も含めれば六つ)の観測位置を設定し、間投詞の出現位置の特徴を調べた。観測位置は、間投詞の存在を無視した文に対して定義し、そこに間投詞が挿入されたかどうかを検査した。

- A: 文頭
- B: 節の前後
- C: 接続詞の前後

D: 節の 1 文節目の直後

ここで、節とは、次のようなものとした。

1. 主部・述部と考えられるものの組みを持つ、文節の集合。
2. ただし、その主部・述部の組みを除いても、残りの文の理解が困難にならないもの。
3. 文も節の一つとし、文の終了も節の終了とする。これは一般に言われる“従属節”等の“節”と同等である。ただし、今回は、連体修飾を行う連体節は節とはしなかった。これは、上記の二つめの条件に触れると考えたためである。なお、書き起こし方の問題でもあるが、間投詞のみで 1 発話となっているようなもの(文頭に現れてはいるが、実質的な文も節も続いている)もあり、そのため節の前後がすべての文頭と一致するわけではない。

以下に、上記観測位置の例を、間投詞を“[]”でくくると共に A~D の記号を付して示す。なお、A~D の複数にあたる場合は対応するものをすべて示す。

- ・ [えと] 学力検査で編入されるわけですが。_{AB}
- ・ 豊橋技術大は豊橋市内にありますから、[え] 豊橋市の南部に位置します。_B
- ・ 分かりました [えと] それでは [あの] お客様のお名前と…_{AB BC}
- ・ お一人 [ま] 2 万円くらいからですかねえ、ありますよ。_{AB D B}
- ・ [あ] (間投詞のみで 1 発話として書き起こされており、文頭ではあるが節の前後ではない場合) これらの観測位置で観測された間投詞数は、表-7 のようになった。

表-8 に、上記観測位置を組み合わせた場合に観測された間投詞数を示す。ここで、“文頭”は“節の前後”に、また“接続詞の前後”も“節の前後”にかなりの割合で含まれるため、3 通りの組み合わせでの結果を示す。

今回対象にした観測位置(“節の前後”等)の総数は、文節の前後という間投詞の現れうると考えられる場所の総数の、およそ 60~70% 程度である。このように、間投詞が現れうると考えられる位置の 60~70% をカバーすることで、全体の 80% 以上の間投詞が観測されているということは、間投詞の出現位置は文節の前後で任意に生じるというよりも、むしろ偏りがあるとい

うことを示唆している。

また、頻出間投詞語彙「え」「えと」「あの」「あ」「ま」の、語彙の出現位置ごとの出現数を表-9に示す。なお、表中の()内の数値は、それぞれの語彙が単独で現れている場合の総数との比(%)である。

この表-9から、「あ」が単独で発声される場合の90%が文頭に現れている。「えと」も60%近くが文頭に現れている。また、「あの」「ま」は節の前後(文頭を除く)に40%現れているなど、間投詞の語彙によってその出現位置に偏りがあることが分かる。

3.2 音響的な変動の分析

間投詞は他の概念語よりも発声のゆらぎが多いことが予想される。そこで、HMMを用いた音節認識システムを用いて、一般の単語と間投詞の尤度をもとに調査を行った。音節認識の結果として得られる尤度が低かったり、あるいは一般の単語よりも尤度が広く分布するようであれば、それは発声のゆらぎが大きいと考えられる。

表-7 各観測位置での出現間投詞数

観測位置	出現間投詞数	出現間投詞数 総間投詞数
A	467	39.4 [%]
B	755	63.7 [%]
C	221	18.7 [%]
D	140	11.8 [%]

表-8 観測された間投詞の総数

観測位置	出現間投詞数	出現間投詞数 総間投詞数
A, B, C, D	974	82.2 [%]
B, C, D	970	81.9 [%]
B, D	883	74.5 [%]

表-9 間投詞語彙による出現位置 (括弧内は%)

間投詞	文頭	節の前後	節の前後 (除く文頭)	接続詞の前後	節の第1文節の直後
え	64 (26.9)	138 (58.0)	89 (37.4)	50 (21.0)	29 (12.2)
えと	133 (57.8)	162 (70.4)	59 (25.6)	51 (22.2)	16 (7.0)
あの	13 (6.3)	101 (48.8)	88 (42.5)	47 (22.7)	43 (20.8)
あ	147 (90.7)	141 (87.0)	9 (5.6)	16 (9.9)	0 (0.0)
ま	11 (8.5)	61 (46.9)	52 (40.0)	15 (11.5)	34 (26.2)
えとですね	25 (75.8)	27 (81.8)	6 (18.2)	5 (15.2)	0 (0.0)

実験では、3人の話者の音声から、頻出間投詞である「え」「えと」「あの」「あ」「ま」という五つの語彙をそれぞれ10個ずつ、また、間投詞以外の普通の単語(平均音節長102ms)を30個ずつ切り出し、不特定話者用の音節単位のHMMを連結したもので照合し、それぞれの尤度を求め、比較した。それぞれの話者における、間投詞と一般単語での尤度の平均と分散を表-10に示す。

表-10から分かるように、間投詞の方が一般の単語よりも尤度の平均は小さく、かつ分散が大きいことから、間投詞の発声のゆらぎは一般の単語に比べて大きいと考えられる。しかし、それほど顕著な差はなく、通常の語彙のように間投詞を音節系列で単語登録することによって音声認識システムで対処できる可能性がある。実験結果を次節で示す。

3.3 音声認識システムによる間投詞登録法の評価

本評価実験では、文脈自由文法のトップダウン型のパーザに基づく連続音声認識システムを用いる¹¹⁾。対象とする文は、「富士山観光案内」タスクの文で、語彙数は約500である。

また、評価用音声データは話者1人の発声した、1文につき一つの間投詞を含むものを50文と、間投詞を含まないものを50文である(認識は、話者適応化モードで行った)。

間投詞を含んだ文の認識のための文法には、新たに、「えと」、「えーと」
「あの」、「あの一」

表-10 間投詞と一般単語の尤度 (log P)

話者	間投詞		一般の単語	
	平均	分散	平均	分散
YM	-41.5	6.93	-40.9	0.88
SK	-44.1	7.78	-42.1	5.37
MY	-42.5	4.51	-41.6	2.07

表-11 評価用 50 文の単語パープレキシティ

評価文	文法の種類		
	文法 0	文法 1	文法 2
間投詞なし	21.6	39.3	24.9
間節詞入り	—	44.7	30.0

表-12 文認識率 [%]
最大ビーム幅: 45 話者: AK

評価文	文法の種類		
	文法 0	文法 1	文法 2
間投詞なし	86.0	82.0 (70.0)	88.0 (76.0)
間節詞入り	6.0	72.0 (28.0)	74.0 (34.0)

ただし、() 内は間投詞も正しく認識された認識率

「え」、「えー」

「あ」、「あー」

「ま」、「まー」

の計 10 個の間投詞を語彙として追加登録した。更に、これらの間投詞が文頭などに入ることができるように書き換え規則を修正した。

間投詞を含んだ文法は、文中の任意の単語の間に出現可能とした文法(文法 1)と、文頭及び助詞の後ろに間投詞を認めるようにした文法(文法 2)との、二つを用意した。また、間投詞を認めないオリジナルな文法(文法 0)も用意した。前述した評価用データ 50 文における、これらの 3 通りの文法についてのテストセットパープレキシティを表-11 に示す。文認識結果を次の表-12 に示す。

この結果から、間投詞を 10 個に絞った認識では、これらを単語登録するという方法は、本タスクの音声認識システムとして十分良い認識精度が得られることが分かった。

4. ポーズの特徴分析

4.1 ポーズ長の分布

ここでは、間投詞や言い直しと、ポーズとの関連を中心に調査を行った。まず表-13 に、10 ms 以上の無音区間をポーズと見なしたポーズ長の分布を示す。なお、今回は、促音による無音区間及び破裂音の直前の無音区間はポーズと見なさず、対象からはずしてある。

4.2 間投詞とポーズ

間投詞の直前と直後に現れるポーズ長の分布を表-14 に示す。なお、A は文頭/文末の間投詞を含めての割

表-13 ポーズ長の分布

ポーズ長	個数	[%]
10-100 ms	464	25.6
100-200 ms	224	12.4
200-300 ms	156	8.6
300-400 ms	132	7.3
400-500 ms	125	6.9
500-600 ms	137	7.6
600-700 ms	96	5.3
700-800 ms	71	3.9
800-900 ms	73	4.0
900-1,000 ms	62	3.4
1,000 ms+	269	14.9
ポーズ総数	1,809 個	

表-14 間投詞直前直後のポーズ長の分布

ポーズ長	直 前			直 後		
	個数	A[%]	B[%]	個数	A[%]	B[%]
ポーズなし	255	21.7	36.9	567	62.3	62.9
ポーズあり	455	38.6	63.1	343	36.8	37.1
文頭/文末	467	39.7	—	8	0.9	—
間投詞総数	1,177 個			910 個		

合; B は文頭/文末の間投詞を除いての割合である。

ここで、間投詞の直前に関しては、“ポーズなし”が A で 22%、B でも 37% となっていることから、ポーズが現れ易い傾向があると思われる。また、逆に、間投詞の直後に関しては、“ポーズなし”が A で 62%、B で 63% となっており、間投詞の直前とは逆に、ポーズが現れ難い傾向があると思われる。また、間投詞の直前直後のいずれかにポーズが現れる割合は 81% 程度になっている。表-14 より、間投詞の直前直後のポーズが独立に現れると仮定した場合、どちらかにポーズが現れる割合は 77% ($=1.0-0.36 \times 0.63$) となるので、ポーズは間投詞の前後独立に現れると言える。

4.3 言い直しとポーズ

言い直し対象の直前直後に現れるポーズを表-15 に示す。ここで、言い直し対象とは、「(しゅ) 宿舎」という場合では「(しゅ)」の部分を使う。なお、A は文頭の言い直し対象を含めての割合、B は除いての割合である。

ここで、言い直し対象の直前に関しては“ポーズなし”が A で 60%、B で 61% となっていることから、どちらかと言うと言い直し対象の直前にはポーズが現れ難い傾向があると思われる。また、逆に、言い直し対

表-15 言い直し対象直前直後のポーズ長の分布

ポーズ長	直 前			直 後		
	個数	A[%]	B[%]	個数	A[%]	B[%]
ポーズなし	75	60.0	61.5	20	15.5	15.5
ポーズあり	47	37.6	38.5	109	84.5	84.5
文頭	3	2.4	—	0	0.0	—
言い直し総数	125個			129個		

象の直後に関しては、“ポーズあり”が85%となっており、このことから、言い直し対象の直前とは逆に、ポーズが現れ易い傾向があると思われる。また、言い直し対象の直前直後のいずれかにポーズが現れる割合は91%程度になっている。表-15より、言い直し対象の直前直後のポーズが独立に現れると仮定した場合、どちらかにポーズが現れる割合は90% (=1.0-0.15×0.61) となるので、ポーズは言い直し対象の前後で独立に現れると言える。

4.4 文節内、文節境界に現れるポーズ

一般に、文節内にはポーズは現れないと言われているが¹⁰⁾、今回対象とした対話においては、文節内に計231個のポーズが現れていた。これは、今回対象としたポーズの13%にもものぼるものである。まず、文節内と文節の境界に現れていたポーズ長の分布を表-16に示す。表-16より、文節境界に現れるポーズ長の分布は、ポーズ全体のものに似ているが、文節内に現れるものは、100ms未満が68%と、かなり多くなっている。

また、今回対象とした対話に含まれる文節数は、4,597個であるが、少なくともこの5%の文節内にはポーズが現れていることになる。これは少数ではあるが、無視はできない程度の割合である。そこで、文節内において、どのような場所にポーズが現れているのかを表-17に示す。

終助詞や格助詞、接続助詞の直前に現れるポーズが計131個、57%と半数以上を占めている。また、複合語中の単語と考えられるものの境界に現れているものが24個、10%を占めている。文節内で行われている言い直しに伴うポーズが10個あったが、これらはいずれも100ms以上のポーズであった。単語内においてポーズが現れているものも28個あった。ただし、28個のうち、24個が80ms未満と、かなり短いものであった。ここで、単語内に現れているポーズと、文節内で行われている言い直しに伴うポーズを除く計193個、84%が単語境界もしくはそれに類する位置に現れている。

表-16 文節内、文節境界のポーズ長

ポーズ長	文節内		文節境界	
	個数	[%]	個数	[%]
10- 100 ms	157	68.0	257	18.4
100- 200 ms	16	6.9	194	13.9
200- 300 ms	16	6.9	125	9.0
300- 400 ms	8	3.5	116	8.3
400- 500 ms	9	3.9	109	7.8
500- 600 ms	3	1.3	125	9.0
600- 700 ms	2	0.9	87	6.2
700- 800 ms	6	2.6	61	4.4
800- 900 ms	3	1.3	64	4.6
900-1,000 ms	2	0.9	50	3.6
1,000 ms+	9	3.9	207	14.8
総数	231個		1,395個	

表-17 文節内のポーズ

内容	100 ms 未満	100 ms 以上	計
	個数	個数	
終助詞「ね」の前	78	1	79
終助詞「よ」の前	4	0	4
格助詞の前	12	22	34
接続助詞の前	14	0	14
複合語の中	10	14	24
名刺接尾辞の前	2	9	11
言い直し	0	10	10
助動詞の前	7	1	8
数詞の間	1	5	6
他の単語境界	5	8	13
単語内	24	4	28
総数	157	74	231

5. 言い直しの特徴分析

5.1 文長と言い直し出現の相関

言い直しは、表-1で見たように、平均するとおよそ10文に1回以上の割合で現れている。図-1から分かるように、およそ文が長くなるほど、言い直しも発生し易いと言える（注：ただし、文長の長いものほどデータが少ないので誤差は大きい）。ここで、発生頻度は、図中の直線で近似できる。すなわち、1文節当たり3%の割合で発生すると言える。このことから、音声対話システムでは言い直しに対処できなければならないことが分かる。

5.2 言い直しの形態

言い直しの形態についての結果を表-18に示す。なお、表中の記法であるが、“R”は同一語句の繰り返し、

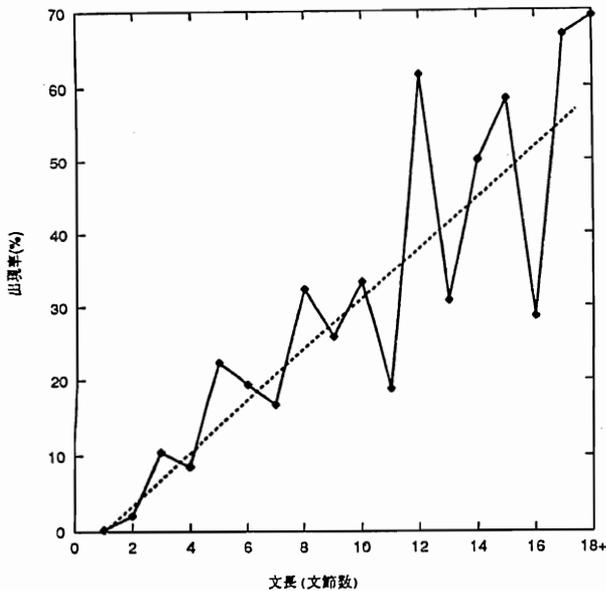


図-1 文長と言い直し出現率

表-18 言い直しの形態

形態	個数
R1:1 R1:1 どちらかを受けて(いた)いただきます。	47
S1:1 S1:1 情報工学課程を(受け)受験されるとなりますと、	27
R1:1 S1:1 R1:1 S1:1 ほとんど(並行に)並行でやってもんですから、	9
R1:1 Is1:2 R1:1 たとえば、(一の)滋賀の一の瀬の方ですねえ。	8
単語内(単語を繰り返さない) ホテルの方がそうな(る)りますね。	8
D1:1 (希望,)試験とかはあるんでしょうか。	6
S1:1 R1:1 S1:1 R1:1 一応(大学院への)修士課程への全員進学が原則 です。	4
R1:2 R1:2 おそらく(一人ず)一人ずつの面接に	4
その他	40
計	153

“S”は語句が同義もしくは同じ文法カテゴリのものに変更されていることを示す。また、“Is”は挿入が行われていること、“D”は語句の切り捨て、つまり言い直しによって文の内容が当初のものから変更されていることを表す。これら“R,”“S,”“Is,”“D”に続く数字は、それらの対応関係を示し、また“:”の後にある

表-19 言い直し対象の単語数

対象単語数	出現数	[%]
0	8	5.6 %
1	92	64.3 %
2	30	21.0 %
3	5	3.5 %
4	5	3.5 %
5	2	1.4 %
6	1	0.7 %

数字はそれぞれの単語数を表す。また、例文の“()”でくくられている部分を言い直し対象、その後の部分を言い直しと呼ぶ。なお、表が複雑になるため、言い直し対象の直後などの間投詞は無視しての集計としてある。表-18に示されるように、1単語の単純な繰り返し、もしくは置換が他のものに比べるとかなり多く出現しているものの、大多数を占めるような形態はない。これは、言い直しが話者のフィードバックによって行われると考えれば、納得のいく結果である。

また、言い直し対象となっている部分の構成単語数を表-19に示す。この結果は、文献7)に示される英語における結果とほぼ同じ結果になっており、言い直し対象となっている単語が1~2単語の場合で全体の85%と、かなりの部分を占めている。なお、ここで、対象単語数0というのは、次の例のように、単語内での発声間違いにおいて単語を言い直すことなく間違えた音からそのまま正しい音を発声するというものである。

「ホテルの方がそうな(る)りますよね。」

また、今回、言い直し対象の直後に、44回(言い直し総数の29%)間投詞が現れていた。このうち、「え」が36.4%、「あ」が20.4%、「あの」が18.2%と上位三つで75%を占めていた。また、言い直しの形態との関係では、“R1:1 R1:1”に間投詞が現れたのが8回と、形態とは無関係に間投詞が現れると仮定した場合(47×0.29=14回)の6割程度の出現数であったこと、また“単語内”の場合には間投詞が現れなかったこと(形態と無関係と仮定した場合には2回)、逆に“R1:1 S1:1 R1:1 S1:1”や“D1:1”では6回(形態と無関係と仮定した場合にはそれぞれ3回と2回)も間投詞が現れたことなどから、言い直し対象に間投詞が続くかどうかは言い直しの形態と何等かの関連があると思われるが、結論づけるには資料が少ない。

5.3 音響的な変動の解析

2音節以上の長さの、“R R”タイプ(30個)の言い

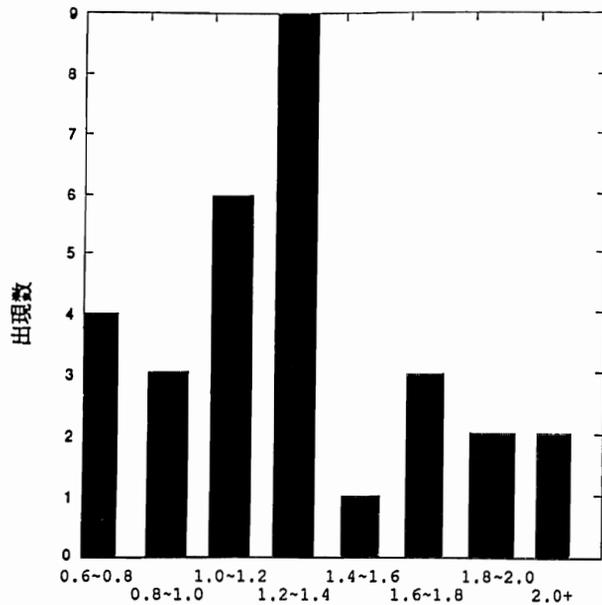


図-2 言い直し対象区間長と言い直し区間長の比

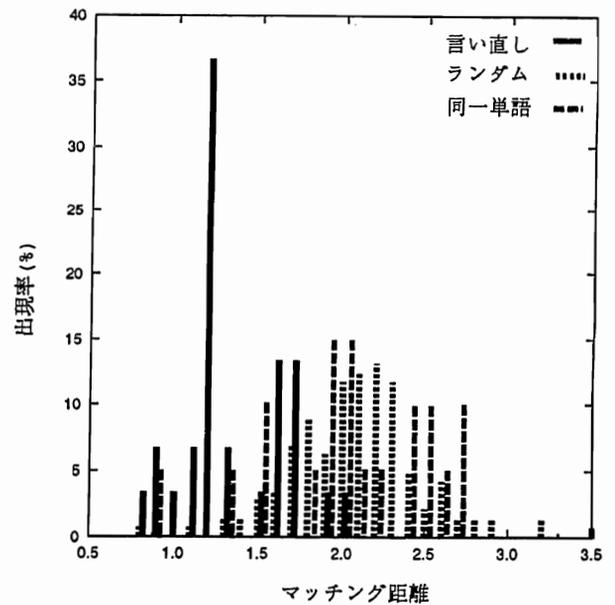


図-3 マッチング距離の分布

表-20 マッチング距離

	平均フレーム数	平均	最大	最小	分散
言い直し対象と 言い直し区間	37.7	1.39	2.06	0.80	0.10
ランダムな2区間	41.0	2.14	3.52	0.84	0.16
同一単語の2区間	50.5	2.09	2.77	0.92	0.24

直しでの、言い直しの対象となる部分と、言い直している部分との長さの比(言い直し対象区間の長さ/言い直し区間の長さ)を、図-2 に示す。30 個中、28 個が0.6 以上 2.0 未満に入っており、特に 1.0 以上 1.4 未満に半数が入っていることから、言い直しに際して発話時間が大きく異なることにはならないが、若干短めになる傾向がある。

言い直し対象区間と言い直し区間のマッチングの平均距離 (30 対) と、ランダムに抜き出した音声データ同士 (145 対) の DP マッチングを行った場合、また同一単語を発声した場合の音声データ同士 (20 対) のマッチングの平均距離を表-20 に示す。なお、同一単語の場合は、文献 13) で用いた音声データを使用している結果である。

なお、「ランダム」とは、音声データからオーバーラップがないように、ランダムに抽出した 30~50 フレーム長の音声データを、同一話者のデータ同士でマッチングを行った結果である。また、図-3 に、各々の分布を示す。平均としてみれば、言い直し区間の平均の方が小さい

が、言い直し区間での平均距離と、ランダムなマッチングでの平均距離とは重なっている部分も多い。言い直しの場合であれば、距離が 1.8 以上のものは 7% 程度と少ないが、ランダムの場合でも、距離が 1.8 以下の場合が 27% 程度とかなりの割合になる。また、同一単語の言い直しの場合でも、言い直しの方が平均距離が小さくなっているため、言い直し区間の自動検出ができる可能性がある。これは次節で検討する。

5.4 言い直しの自動検出の試み

本実験では、音声データのみを用いて、DP マッチングによるワードスポッティング法によって、「R R」という形態の言い直しの自動検出を試みた。

方法としては、無音区間を取り除いた音声データの言い直し対象区間と言い直し区間は連続しているため、その境界をあるフレーム (JS) と仮定する。そして、JS の前の何フレームかと後の何フレームかに対して DP マッチングを行う (フレーム数は可変)。ただし、音声データ中のどこに言い直しが存在するかは分からないため、JS は音声データの始点から終点まで、1 回のマッチングごとに 1 フレームずつ移動する。最終的に、マッチングの平均距離の小さいパスが、言い直しの候補であると考えられる。

このような方法によって、自動検出を試みた結果では、言い直し対象区間の始点と、言い直し区間の始点のみに注目した場合、得られた候補の上位 10 個を考えても、全体の 40% 程度しか検出できなかった。これは、およそ 1 音節の 1/3 程度のズレを認めた場合であるが、仮に 1 音節の 2/3 程度のズレを認めても、上位 10

個を考えても 60%程度が正解になるにすぎなかった。また、4.3 節の結果から、パワーの谷を利用して、言い直し対象の終点の候補を得る方法も試みたが、良い結果は得られなかった。

これは、前節で述べたように、マッチングの平均距離だけでは言い直し以外の箇所と重なる部分もあることが一つの理由として考えられる。また、言い直し区間の一部であるならば、始点や終点が正解からかなりズレてはいるものの、かなりの割合で検出できている。そのため、仮に本来検出したい部分のパスが得られたとしても、その順位が低くなってしまっている場合もあった。これは、対象区間が短いことにもよる。実際の言い直し区間の直前には、ポーズが 84%存在するが、本手法ではこれを利用しなかった。この利用を含め、本手法は今後の改良を要する。

6. ま と め

間投詞の語彙としては、少数の語彙が多用されており、また出現位置にも偏りが見られた。また、間投詞の 63%は直前にポーズを伴っていた。

間投詞を含む文の音声認識による実験では、間投詞の発声のゆらぎに関しては、他の概念語よりもゆらぎが多いものの、その差は小さく、出現頻度の多い間投詞を単語として登録する方法によって、評価実験を行った。この方法は、単語パープレキシティの増加と計算量の増加が問題となるが、文認識の性能としては良好な結果が得られた。計算量の問題に関しては、高頻出間投詞のみを辞書に登録し、その他の間投詞が発声された場合に関しては未知語が挿入されているとして対処する方法⁶⁾によって、かなり緩和される。

一般に文節内にはポーズは現れないと言われていたが、今回は文節内に 231 個のポーズが現れていた。これは、今回対象にした全文節数 (4,597 文節) の 5%程度にあたり、自然な音声対話においては、文節内にもポーズが無視できない割合で現れうるということが分かった。特に、そのうちの 57%は助詞の直前であり、その出現位置がかなり偏っていることが分かった。また、文節内に現れるポーズの 68%が 100 ms 未満と、文節内に現れるポーズは、他の部分に現れるポーズよりも短いものが多かった。

また、言い直しに関しては、その形態の分類は得られたものの、間投詞の場合ほどに顕著な形態というものはなかった。言い直しは、誤りの訂正や、文の構造や単語を変更するために行われているもので、これらの訂正・変更は発話からのフィードバックによって行

われると考えられるため、これは自然なことであろう。言い直しに対しては、より高度な、文法などを考慮した分析を更に進める必要がある。また、言い直し対象の 84%は直後にポーズを伴っていた。

更に、言い直し対象と言い直しは、音響的には似ているものの、言い直し以外でも同程度に類似した部分が現れる場合も少なくない。また、始点・終点では音声パターンの曖昧さのためにマッチングがうまくいかないなどのことから、自動検出は良い結果が得られなかった。ポーズの利用などを含め、今後の検討を要する。

文 献

- 1) 有田英一, 小暮 潔, 野垣内出, 飯田 仁, “メディアに依存する会話の様式—電話会話とキーボード会話の比較—,” 情報処理学会研究報告 87-NL-61-5 (1987).
- 2) 山本幹雄, 小林 聡, 中川聖一, “音声対話文における助詞落ち・倒置の分析とその解析手法,” 情報処理学会論文誌 33, 1322-1330 (1992).
- 3) 小林 聡, 甲斐充彦, 山本幹雄, 中川聖一, “間投詞の出現位置の特徴分析と音声認識システムの評価,” 信学会第 2 種研資 SPREC-92-3, 21-25 (1993).
- 4) 高木一幸, 保浦直子, 板橋秀一, “模擬対話音声における各種区分の持続時間の性質,” 信学技法 SP 92-111, 63-70 (1992).
- 5) 高木一幸, 保浦直子, 板橋秀一, “対話音声の中の発話単位の時間関係,” 音講論集 3-8-2 (1993.3).
- 6) 甲斐充彦, 中川聖一, “連続音声認識システム SPOUJS-SYNO における間投詞, 未知語処理の検討,” 音講論集 2-9-14 (1993.10).
- 7) E. Shriberg, J. Bear and J. Dowding, “Automatic detection and correction of repairs in human-computer dialog,” Proc. Speech and Natural Language Workshop, 419-424 (1992).
- 8) D. O’Shaughnessy, “Analysis of false starts in spontaneous speech,” Proc. Int. Conf. Spoken Language Processing, 931-934 (1992).
- 9) C. Nakatani and J. Hirschberg, “A corpus-based study of repair cues in spontaneous speech,” J. Acoust. Soc. Am. 95, 1603-1616 (1994).
- 10) 橋本進吉, 日本の言語学 第 3 巻 文法 I, 服部四郎, 大野 晋, 阪倉篤義, 松村 明編 (大修館書店, 東京, 1978), pp. 439-443.
- 11) 甲斐充彦, 中川聖一, “ワードスポッティング法を用いた文脈自由文法制御フレーム同期型 HMM 連続音声認識法,” 信学論 J 76-D-II, 1329-1336 (1993).
- 12) 保坂順子, 衛藤純司, “話しことばにおけるポーズ節の考察,” 情報処理学会第 48 会 (平成 6 年前期) 全国大会講論集 2-Q-6 (1994.3).
- 13) 森屋裕治, 阿部野尚, 山本幹雄, 中川聖一, “対話予測を利用した音声による観光案内対話システム,” 信学技法 SP 92-121 (1993.1).
- 14) 武田一哉, 小沼知浩, “自由発話文理解のための Garbage HMM の利用の検討,” 信学技法 SP 92-127 (1993.1).
- 15) 佐川雄二, 大西 昇, 杉江 昇, “自己修復を含む日本語不適格文分析とその計算機による理解手法に関する考察,” 情報処理学会論文誌 35, 46-52 (1994).